REMARKS

The present application was filed on October 11, 2000 with claims 1-30. Claims 1-30 remain pending. Claims 1, 10, 11, 20, 21 and 30 are the pending independent claims.

In the outstanding final Office Action dated March 8, 2004, the Examiner: (i) rejected claims 1-3, 6-9, 11-13, 16-19, 21-23 and 26-29 under 35 U.S.C. §102(a) as being anticipated by Knorr et al., "Distance-Based Outliers: Algorithms and Applications," (hereinafter "Knorr"); and (ii) rejected claims 10, 20 and 30 under 35 U.S.C. §102(a) as being anticipated by Sheikholeslami et al., "WaveCluster: A Wavelet-Based Clustering Approach for Spatial Data in Very Large Databases," (hereinafter "Sheikholeslami").

Applicants acknowledge the indication of allowable subject matter in claims 4, 5, 14, 15, 24 and 25.

With regard to the rejection of claims 1-3, 6-9, 11-13, 16-19, 21-23 and 26-29 under 35 U.S.C. §102(a) as being anticipated by Knorr, Applicants assert that such claims are patentable for the following reasons. Independent claims 1, 11 and 21 recite techniques for detecting one or more outliers in a data set. One or more sets of dimensions and corresponding ranges in the data set, which are sparse in density, are determined. One or more data points in these sets are identified as the outliers in the data set.

The present invention identifies outliers by observing the density distributions of projections from the data. It considers a point to be an outlier if, in some lower dimensional projection, it is present in a local region of abnormally low density. More specifically, the present invention defines outliers in the data set by examining at those projections of the data having an abnormally low density. By defining clusters which are specific to particular projections of the data, it is possible to design more effective techniques for finding clusters.

Knorr discloses algorithms and applications for finding outliers using a distance based approach, and teaches that outliers are found by answering a nearest neighbor or range query with a specified radius for each object. If more than a specified number of neighbors are found within the range, or in the neighborhood of the object, it is declared a non-outlier. The object is declared an outlier if the number of neighbors found in the range is less than or equal to the specified number.

Knorr suffers from the inherent disadvantage of treating the data in a uniform way even though different localities of the data may contain clusters of varying density. When finding the outliers based on the density of their local neighborhoods and defining distances in full dimensional space, all pairs of points are almost equidistant and it becomes difficult to use these measures effectively.

Knorr focuses on finding outliers in multidimensional data sets, for example, "k-dimensional data sets with large values of k (e.g. $k \geq 5$)," (Abstract). However, Knorr does not focus on the high dimensionality aspect of outlier detection, involving dimensionalities of 100 or 200, as in the present invention. Therefore, Knorr uses methods which are more applicable for low dimensional problems, such as relatively straightforward proximity measures of which the complexity increases exponentially with dimensionality. Thus, for relatively smaller dimensionalities of 8 to 10, the technique of Knorr is computationally intensive. For higher dimensionalities, the technique is likely to be infeasible from a computational standpoint.

Regarding independent claims 1, 11 and 21, Knorr fails to disclose a technique for determining sets of dimensions and ranges in the data set which are sparse in density. The Examiner refers to the Abstract and paragraphs one and two of section 3.1 of Knorr in rejecting this element of independent claims 1, 11 and 21. However, a defined radius range query performed for each object does not provide the support necessary for an anticipation rejection since it differs significantly from a determination of one or more sets of dimensions and corresponding ranges which are sparse in density. This element of independent claims 1, 11 and 21 is not addressed in the Examiner's response to Applicant's previous arguments. As described above, while Knorr simply determines whether an object is an outlier by the number of neighbors found within a specified range, independent claims 1, 11 and 21 of the present invention recite the determination of projections (dimensions and corresponding ranges) in the data having an abnormally low density of objects (sparse in density).

Further, Knorr fails to disclose the identification of data points in the sets of dimensions and ranges as outliers. In response to Applicants previous arguments, the Examiner states that Knorr clearly teaches the identification of data in sets of dimensions and ranges. However, the Examiner

fails to realize that since Knorr fails to disclose the determination of sets of dimensions and corresponding ranges in the data set which are sparse in density, it also fails to disclose the identification of data points in these sets of dimensions and ranges as outliers.

Applicants assert that dependent claims 2, 3, 6-9, 12, 13, 16-19, 22, 23 and 26-29 are patentable for at least the reasons that independent claims 1, 11 and 21, from which they depend, are patentable. Further, dependent claims 2, 3, 6-9, 12, 13, 16-19, 22, 23 and 26-29 recite patentable subject matter in their own right. Accordingly, withdrawal of the rejection to claims 1-3, 6-9, 11-13, 16-19, 21-23 and 26-29 under 35 U.S.C. §102(a) is therefore respectfully requested.

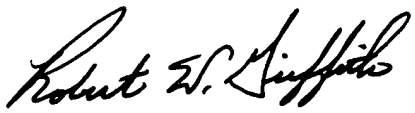
With regard to the rejection of claims 10, 20 and 30 under 35 U.S.C. §102(a) as being anticipated by Sheikholeslami, Applicants contend that such claims are patentable for the following reasons. The techniques of claims 10, 20 and 30 recite the detection of one or more outliers in a data set. One or more patterns in the data set are identified and mined which have abnormally low presence not due to randomness, and one or more records having the patterns present in them are identified as outliers.

Sheikholeslami discloses a wavelet-based clustering approach for spatial data in very large databases. More specifically, Sheikholeslami describes the discarding of noise objects (outliers) during the mining process in stating that it is "insensitive to noise" (Abstract), thereby teaching away from independent claims 10, 20 and 30 of the present invention. Therefore, while Sheikholeslami describes a clustering algorithm that is able to identify clusters irrespective of their shapes or relative positions, it fails to disclose the identification of patterns in the data set which have abnormally low presence not due to randomness.

Further, since Sheikholeslami fails to disclose the identification of such patterns having abnormally low presence, it also fails to disclose the identification of records as outliers that have the patterns present. Accordingly, withdrawal of the rejection to claims 10, 20 and 30 under 35 U.S.C. §102(a) is therefore respectfully requested.

In view of the above, Applicants believe that claims 1-30 are in condition for allowance, and respectfully request withdrawal of the §102(a) rejections.

Respectfully submitted,



Date: June 8, 2004

Robert W. Griffith
Attorney for Applicant(s)
Reg. No. 48,956
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-4547